A Filter Bank Approach To Earthquake Early Warning

Men-Andrin Meier[1], Tom Heaton[2] and John Clinton[1]

[1] Swiss Seismological Service, ETH Zürich, [2] Caltech Seismological Laboratory

Contact: menandrin@gmail.com

i. Introduction

Earthquake Early Warning (EEW) is a race against time. The longer it takes to detect and characterize an ongoing event, the larger is the blind zone - the region where a warning arrives only after the most damaging ground motion has occurred. The problem is most acute during destructive medium size earthquakes, where damaging ground motion is confined to a small zone around the epicenter. An ideal EEW algorithm which is fast enough to reliably provide relevant alerts for such scenario events would have to exploit available real-time information in a more optimal way than what is currently done by existing algorithms. In this study we present a novel approach to EEW which fully mines the broadband frequency content of incoming waveforms. We extend the Virtual Seismologist method of *Cua and Heaton 2007* to an evolutionary EEW algorithm that starts parameter estimations at the p-wave onset on the first station. We use a filter bank with minimum phase delay filters which allows us to use frequency information from each frequency band at each triggered station at the earliest possible time. With an extensive dataset of near-field earthquake waveforms we demonstrate the potential of such a processing scheme to infer earthquake source parameters in real-time with high accuracy, starting from observations at a single station.

ii. Data & Filter Bank

We measure the frequency content of seismic waveforms in real-time with the goal of inferring the two source parameters *magnitude* and *hypocentral distance*. We have compiled an extensive near-field record waveform data set with three component records including

- broadband and strong motion data from southern California (SCSN)
- strong motion data from Japan (kNet & kikNet)
- digital NGA West1 data

We process a total of ~160,000 waveform traces with magnitudes $2 \leq M \leq 7.9$ and hypocentral distances ≤ 100km from shallow crustal earthquakes. We pass all traces through the filter bank, a set of 9 octave wide 4th order Butterworth passband filters between 0.1 - 48Hz (***Figure 1***). This filter bank operation is a class of wavelet transform and it optimally solves the trade-off between time and frequency resolution. Low frequencies are delayed more than high frequencies (***Figure 2***), but the delay is the minimum that is physically possible. This way the information from each frequency band becomes available at the earliest possible point in time.

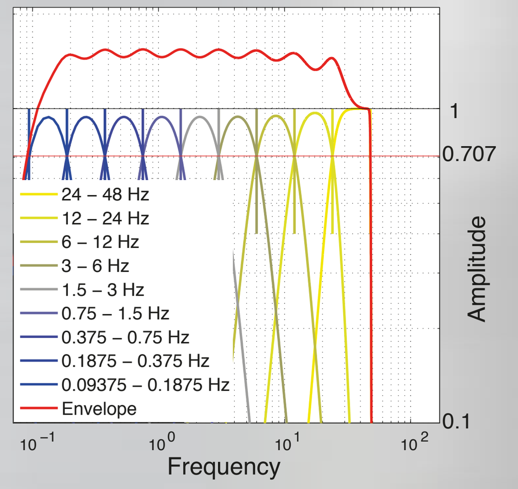

Figure 1: *Frequency response of the nine bandpass filters.*

For each waveform trace of the data set, we obtain nine bandpass filtered seismograms, on each of which we measure peak absolute amplitudes as a function of time since the p-wave onset. These maximum absolute amplitudes we term *narrow band peak ground velocities*, ***PGV_{nb}***.

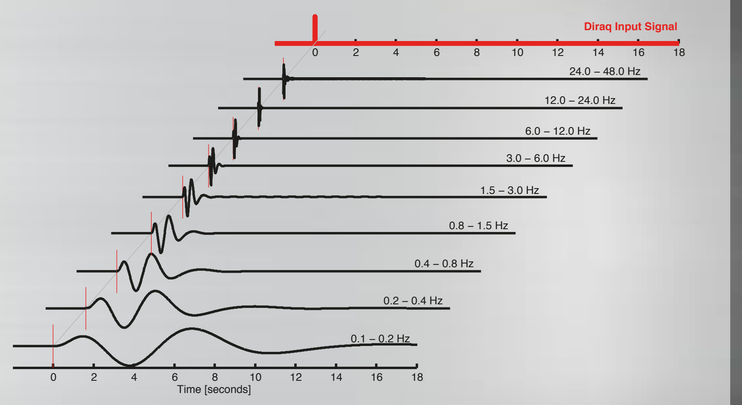

Figure 2: *Impulse response of the nine bandpass filters.*

iii. Parameter Inference

In the following we determine how diagnostic the PGV_{nb} observations are for the two parameters *magnitude* ***m*** and *hypocentral distance* ***r***. We present two different algorithms with which the parameters can be estimated in real-time for any waveform trace. ***Figures 3 and 4*** illustrate the two algorithms with a near-field record of a large event (M6.5 at 11km) and a more distant record of a small event (M3.2 at 31km).

M6.5 at 11km hypocentral distance, Japan

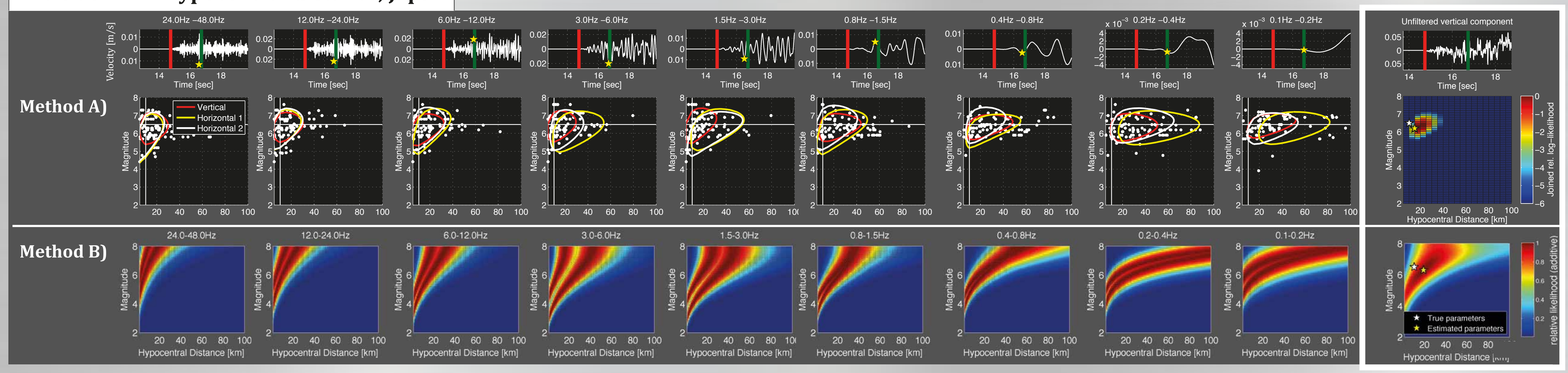

Figure 3: *Example application of methods A and B to a M6.5 record with a hypocentral distance of 11km. Top row: Bandpass filtered seismograms with increasingly low corner frequencies (from left to right). Middle row: 50% contours of relative log-likelihood functions from method A. Bottom row: relative linear likelihood functions of method B. The last figures of the middle and bottom rows show the joined relative likelihood functions from all bands and the inferred maximum likelihood estimates (yellow stars) and catalog values (white stars).*

M3.2 at 31km hypocentral distance, southern California

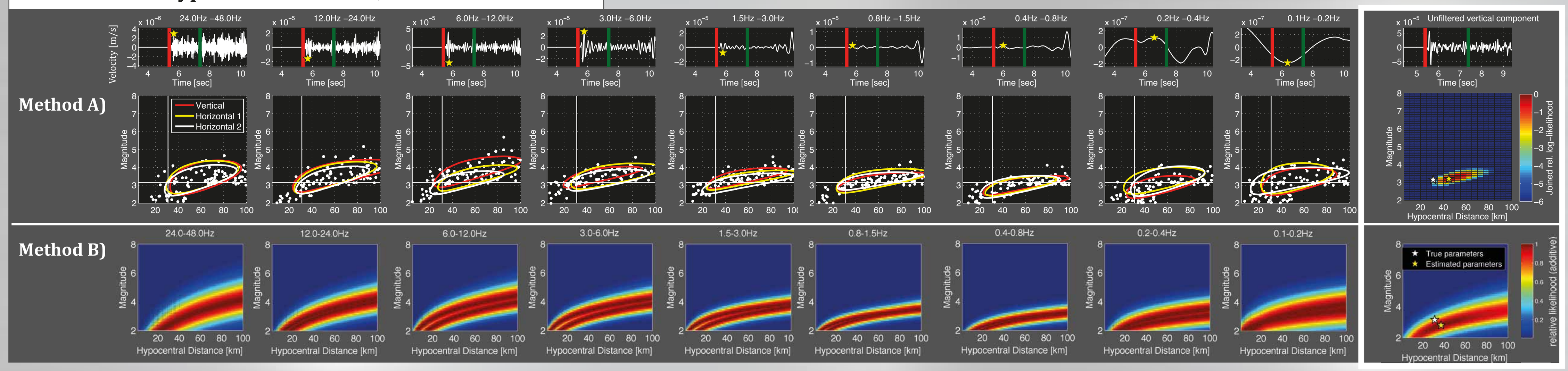

Figure 4: *Same as Figure 4, but for a M3.2 record at 31km hypocentral distance.*

Method A: Non-parametric empirical maximum likelihood

In this memory based approach we measure PGV_{nb} of the target trace and extract from the compiled waveform data set those traces that have produced the most similar amplitudes. This is performed in each frequency band and for each component individually. The magnitude and hypocentral distances of those traces are approximated with a bivariate Gaussian distribution which is taken to be the empirical likelihood function of the two parameters. We then combine the 9 likelihood functions from the different frequency bands and maximise the joint likelihood function to obtain the most likely estimates $\boldsymbol{M_{MLE}}$ and $\boldsymbol{R_{MLE}}$.

Method B: Parametric non-linear regression

We fit a parametric regression model to the PGV_{nb} values in each frequency band. We use a modified version of the Campbell 1981 model

$$g(M,R,\beta) = log[PGV_{nb,i}] = log[a] + b_1 M_i - b_2 M_i^2 - (d_1 + d_2 M_i) log[R_i + c_1]$$

where $\beta = [a, b_1, b_2, c_1, d_1, d_2]^T$ are regression coefficients, M_i, R_i and $PGV_{nb,i}$ are the magnitude, hypocentral distance and observed peak velocities for the i-th trace, respectively. Once the coefficients are determined, the probabilistic amplitudes for any parameter combination can be expressed as a likelihood function:

$$\mathcal{L}(\boldsymbol{M},\boldsymbol{R}|\boldsymbol{PGV_{nb}},\beta,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(log[\mathbf{PGV_{nb}}] - g(\mathbf{M},\mathbf{R},\beta))^T (log[\mathbf{PGV_{nb}}] - g(\mathbf{M},\mathbf{R},\beta))\right)$$

For a set of observed PGV_{nb} values from a target trace, the most likely parameter combination can then be found by maximising the joint likelihood function.

Question to the reader: How are *dependent* likelihood functions optimally combined? The PGV_{nb} values from neighbouring frequency bands are correlated and hence not independent. Can we use a free form mixture model by adding the likelihood functions? Feedback and discussion are welcome.

iv. Results

We have applied both algorithms to 120,000 randomly selected traces from the data base. The magnitude prediction errors obtained when using the first 3 seconds of each trace are shown in ***Figure 5***.

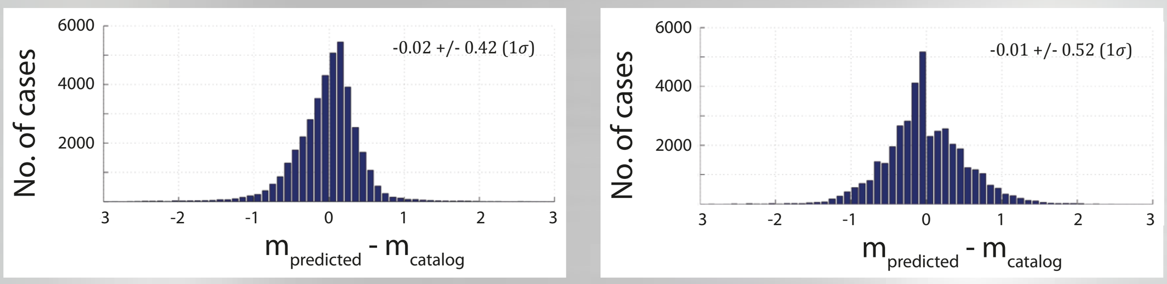

Figure 5: *Magnitude prediction error for method A (left) and method B (right).*

Method A:
- \+ no bias (mean ~ 0) and low scatter around mean ($1\sigma \sim 0.42$), comparable to multi-station estimates of published EEW algorithms
- \- magnitude saturation, underestimation of events with M>7
- \- near-field bias: systematic overestimation of magnitudes of near-field traces, presumably because we do not account for whether the S-phase is already contained in the PGV_{nb} observations or not

Method B:
- \+ does not have the near-field bias that method A has
- \+ once the coefficients are computed, estimations are computationally cheap
- \+ allows predictions of parameters that have not yet been observed, e.g. M>8
- \- larger scatter than method A

v. Conclusions and Outlook

- We have presented a novel approach for inferring magnitudes and hypocentral distances in real time, based on single station records.
- By simultaneously considering multiple frequency bands we can resolve the ambiguity between the two main parameters *magnitude* and *hypocentral distance*.
- With an extensive data set we have demonstrated promising generalisation performances for magnitude and source/station distance prediction: $1\sigma \sim 0.42$ magnitude units and ~ 17 km, respecetively, using the first 3 seconds of real-time waveforms. Such high accuracy is necessary for useful EEW applications in the crucial near-source region.
- Becasue the algorithm is based on partly different data than other EEW algorithms such as ElarmS, OnSite or PRESTo, the filter bank algorithm can potentially provide information gain if used in combination with those algorithms.
- Owing to the the probabilistic formulation it is straight forward to combine the estimates of multiple stations and to include prior information and externally derived location estimates.
- We are currently working on a real-time implementation of the algorithm at the Southern California Seismic Network.
- Future developments will include the real-time recognition of S-phase arrivals, the use of prior information and extending the algorithm to a full multi-station EEW system.